PIMMS tools for capturing metadata about simulations

Charlotte Pascoe (1), Gerard Devine (2), Gregory Tourte (3), Stephen Pascoe (1), Bryan Lawrence (2), and Hannah Barjat (4)

(1) RAL Space, STFC, STFC, United Kingdom (charlotte.pascoe@stfc.ac.uk), (2) University of Reading, United Kingdom, (3) University of Bristol, United Kingdom, (4) Barjat Consulting, United Kingdom

PIMMS (Portable Infrastructure for the Metafor Metadata System) provides a method for consistent and comprehensive documentation of modelling activities that enables the sharing of simulation data and model configuration information. The aim of PIMMS is to package the metadata infrastructure developed by Metafor for CMIP5 so that it can be used by climate modelling groups in UK Universities.

PIMMS tools capture information about simulations from the design of experiments to the implementation of experiments via simulations that run models. PIMMS uses the Metafor methodology which consists of a Common Information Model (CIM), Controlled Vocabularies (CV) and software tools. PIMMS software tools provide for the creation and consumption of CIM content via a web services infrastructure and portal developed by the ES-DOC community. PIMMS metadata integrates with the ESGF data infrastructure via the mapping of vocabularies onto ESGF facets.

There are three paradigms of PIMMS metadata collection:

- Model Intercomparision Projects (MIPs) where a standard set of questions is asked of all models which perform standard sets of experiments.

- Disciplinary level metadata collection where a standard set of questions is asked of all models but experiments are specified by users.

- Bespoke metadata creation where the users define questions about both models and experiments.

Examples will be shown of how PIMMS has been configured to suit each of these three paradigms. In each case PIMMS allows users to provide additional metadata beyond that which is asked for in an initial deployment.

The primary target for PIMMS is the UK climate modelling community where it is common practice to reuse model configurations from other researchers. This culture of collaboration exists in part because climate models are very complex with many variables that can be modified. Therefore it has become common practice to begin a series of experiments by using another climate model configuration as a starting point. Usually this other configuration is provided by a researcher in the same research group or by a previous collaborator with whom there is an existing scientific relationship. Some efforts have been made at the university department level to create documentation but there is a wide diversity in the scope and purpose of this information. The consistent and comprehensive documentation enabled by PIMMS will enable the wider sharing of climate model data and configuration information.

The PIMMS methodology assumes an initial effort to document standard model configurations. Once these descriptions have been created users need only describe the specific way in which their model configuration is different from the standard. Thus the documentation burden on the user is specific to the experiment they are performing and fits easily into the workflow of doing their science.

PIMMS metadata is independent of data and as such is ideally suited for documenting model development. PIMMS provides a framework for sharing information about failed model configurations for which data are not kept, the negative results that don't appear in scientific literature.